SATELLITE-CONSTELLATION DESIGN

This article presents the rudiments of constellation design to familiarize the reader with some of this field's methods, issues, and concerns. Two exotic constellations are under consideration for various missions.

Sputnik's launch in October 1957 propelled the second half of the 20th century into the space age. Since then, space-based observation and communications have dominated artificial satellite use. The use of circular, elliptical, and synchronous orbits, combined with dynamical effects of Earth's J_2 gravitational harmonics, produce an array of constellations with specific properties to support various mission constraints. A key point is that constellation design is not the mere repetition of a satellite orbit as a template.

The aggregate of satellites and orbits presents a totally new systems problem with a combinatorial growth in complexity for the most elementary analysis questions. The fundamental performance metric is constellation geometric-coverage statistics. Ergodic theory, the multiresolution visual calculus, and dynamical systems theory are some new approaches to constellation design problems. New infrastructures and new paradigms will further develop this technology.

The importance of **communications satellite** constellations cannot be overstated. In one fell swoop, such a constellation can provide an underdeveloped region without a modern communication infrastructure with an instant modern communications network. The social and economic implications of this technology are enormous.

MARTIN W. LO
Jet Propulsion Laboratory
California Institute of Technology

The traveling-salesman problem

The need for satellite constellations is a natural outgrowth of the need for global communications and intelligence gathering. A single satellite can only observe a spherical cap region of the Earth (as illustrated in Figure 1), called the instantaneous nadir-pointing coverage circle (also referred to as a footprint) of the satellite. The coverage circle's area, A, is a function of altitude, h, and the Earth's radius, R (6,378.14 km),

$$A(h) = 2\pi R^2 (1 - \cos \theta) = 2\pi R^2 h/(R + h), \quad (1)$$

which is always smaller than the hemisphere's area, $2\pi R^2$. Hence, we can get a quick estimate of the lower bound for the number, N, of satellites with altitude, h, required to provide instantaneous global coverage by dividing the Earth's surface by $A(h)$:

$$N > 4\pi R^2 / A(h) = 2 + 2R/h. \quad (2)$$

Although Equation 2 is not a tight estimate, it tells us right away that we must always have more than two satellites to provide simultaneous global coverage. Thus satellite constellations are essential to provide any kind of global satellite service.

At first glance, it might seem that constellation design is merely the act of replicating multiple copies of a single satellite in slightly different orbits. Although the end result might seem to support that idea, the design process is actually very
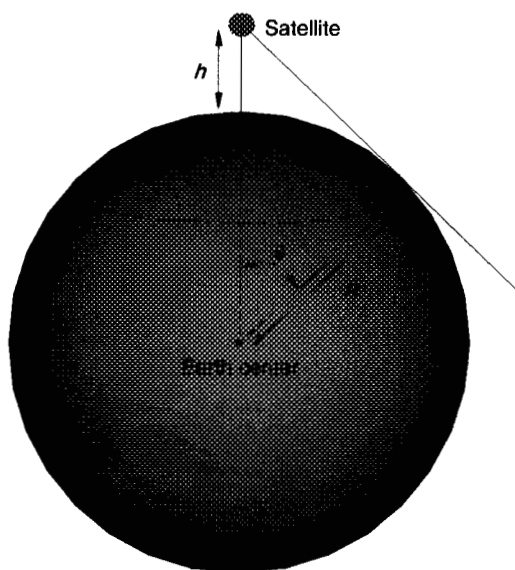
Figure 1. The nadir-pointing coverage circle of a satellite on Earth.

Figure 2. Elliptical orbit in the orbit plane.

difficult. It is a much more organic process somewhat akin to the building of a multicellular organism. In its most basic form, each of the cells is an identical satellite. For more complex missions, satellites with different capabilities, such as differentiated cellular groupings, might be necessary. But even in the simplest case of identical satellites in similar orbits, the constellation performance analysis can be extraordinarily complex due to the combinatorial effects of multiple satellites linking with multiple locations on the ground.

Even before the current interest in constellations, ground-station planners faced the difficult problem of scheduling communications passes for multiple satellites. One way to see why this is such a hard problem is to compare it to the well-known traveling-salesman problem: Given a list of cities, what is the optimal path for a salesman to visit each city once? Now suppose the salesman is a satellite and the path traversed is a fixed orbit. What orbit, if it is even possible, provides the optimum path to visit the list of cities? Now increase the number of satellites, each in a different orbit. What combination of orbits provides the optimum coverage of all ground stations? The once-simple graph-theory problem has not only increased in size and combinatorial complexity, but the paths must now satisfy the constraint of a set of differential equations for artificial satellites. Fortunately for most applications, designers typically use certain well-understood satellite orbits for constellation design.
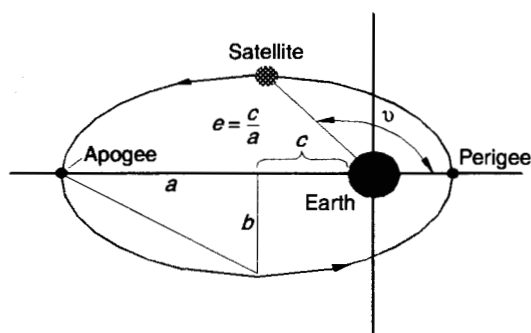
Basic orbital mechanics

To intelligently discuss constellation design, let's cover some basic orbital mechanics to describe the geometry and provide metrics to measure performance. I will touch upon only circular and elliptical orbits, as depicted in Figure 2. The semimajor axis, a, is half the distance of the axis through the ellipse's longest part. The semiminor axis, b, is half the distance of the axis through the ellipse's shortest part. The eccentricity, $e = c/a$, measures the ellipse's shape. For a circular orbit, the eccentricity is 0 because in this case, the axes a and b are equal, and c degenerates to 0. In the limiting case when a approaches infinity with a finite b, c also approaches infinity, and we obtain a parabolic orbit with eccentricity equal to 1. Thus, for elliptical orbits, the eccentricity always falls between 0 and 1. The satellite's location on the elliptical orbit is given by angle v in Figure 2 (called the true anomaly). The closest approach to Earth is called perigee, the farthest is called apogee.

From Keplerian orbit theory, without perturbations, elliptical orbits are always confined to a plane centered on Earth known as the orbit plane. To provide space with coordinates, a fundamental plane must be chosen to represent the xy plane.

For Earth satellites, this plane is usually the equator's plane with the x-axis pointed in the direction of the constellation Aries, also known as the vernal equinox direction. Our orbit plane might be inclined with respect to the equator, and the angle between these two planes is simply called the inclination, usually denoted by i. The orbit plane intersects the equatorial plane in a line known as the *line of nodes*. The line of nodes intersects the orbit in the equatorial plane in two points. The point where the satellite moves into the $+Z$ hemisphere is called the ascending node; the other node is called the descending node. Ω generally denotes the angle between the x-axis and the ascending nodes. Finally, the last para-
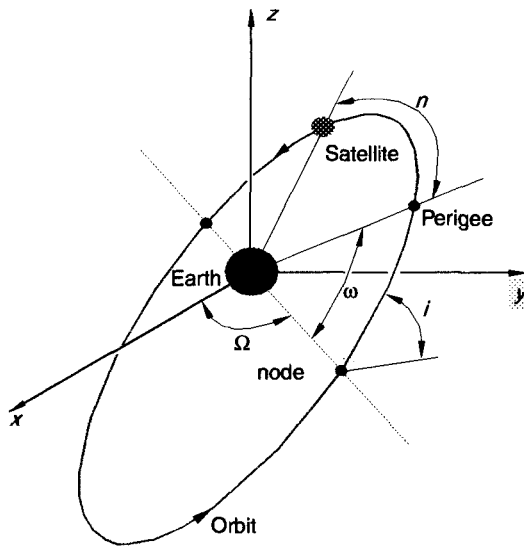
Figure 3. The orbital inclination, ascending node, and the argument of perigee.

meter required to specify the orbit is the angle ω, called the *argument of perigee*, between the ascending node and the perigee. These angles are indicated on the inclined orbit in Figure 3. The six parameters, a, e, i, Ω, ω, and ν—called the classical orbital elements—completely characterize a conic orbit.

Perturbations

For Earth-orbit design, we must consider three major perturbations: atmospheric drag, the gravitation of the sun and moon, and the Earth's equatorial bulge. For typical constellations, average orbit altitude is generally assumed to be high enough to avoid the atmospheric drag that would cause the orbit to decay if unmaintained. Similarly, the altitude is assumed to be no higher than that of geosynchronous orbits (35,863 km) so that luni-solar perturbations are small enough and might be ignored during the design stage. In actual satellite operation, these forces must be modeled and included in the orbit calculations to the degree of accuracy that the mission requires. But during typical constellation design, the main effect that must be included is the so-called J_2 term, named after the coefficient of a planet's geopotential harmonic expansion.

The J_2 term corresponds to the planet's equatorial bulge and has three important effects on the orbit. The equatorial bulge might be thought of as an additional tire of mass around a spherical planet's equator. This extra mass tends to pull the satellite down towards the equatorial plane

sooner than expected. This tends to make the satellite reach its ascending node sooner than it would without the perturbation. Hence the node appears to move backward (see Equation 3) and the effect is called nodal regression. In other words, the orbital plane is precessing due to J_2 perturbation. For a space shuttle to orbit at 200 km altitude and 28.5 degrees (deg) inclination, the nodal regression is roughly -7 deg per day, which is quite significant.

$$d\Omega/dt = -2.06474 \times 10^{14} \, a^{-7/2} \cos(i)$$
$$(1 - e^2)^{-2} \text{ deg/day.} \qquad (3)$$

A second effect due to J_2 perturbation is called the precession of the argument of perigee. J_2 causes the perigee to rotate around the orbit's normal vector, which is also the orbit plane's normal vector (see Equation 4). The orbital period is shortened because as the orbit moves from node to node, it reaches the nodal crossings faster due to the equatorial bulge's gravitational forces:

$$d\omega/dt = 1.03237 \times 10^{14} \, a^{-7/2} (4 - 5 \sin^2(i))$$
$$(1 - e^2)^{-2} \text{ deg/day.} \qquad (4)$$

Orbital zoology

Now that we have the basic tools for orbital mechanics, let's examine some of the space industry's most useful orbits.

Circular orbit. Circular orbits are the most useful because of their symmetry and ease of analysis. They move with uniform speed, and we can frequently estimate their coverage geometry with good, closed-form approximations. However, note that the rate at which circular orbits cover ground is not uniform. In other words, a circular orbit's projected velocity on the ground is not uniform due to the Earth's rotation. All points on Earth rotate at the same angular velocity around the North Pole. However, the actual velocity at the equator is faster than that at a higher latitude. This is because latitude circles have different radii, but the same angular rate. Because velocity is angular rate times radius, the points at different latitude circles have different velocities.

Three distinct circular-orbit classes have emerged in orbital nomenclature: the geosynchronous Earth orbit (GEO), the low Earth orbit (LEO) for orbital altitudes in the 100 to 1,000 km range, and the medium Earth orbit (MEO) for altitudes between GEO and LEO. The cut-off point between LEO and MEO is not, of course, well-defined. A related

acronym for highly elliptical orbit (HEO) refers to orbits with high eccentricity. A final acronym of some use is the geosynchronous transfer orbit (GTO), which has a perigee altitude (usually 200 km) typical of a LEO, and the apogee altitude of a GEO. It is an intermediate orbit used to transfer a satellite from LEO into GEO through the well-known Hohmann transfer depicted in Figure 4.

Geosynchronous orbit. Geosynchronous orbits are perhaps the best known of all orbits due to their extreme usefulness for the communications satellite industry. The idea is as simple as it is elegant: find a circular orbit around the equator that moves at the same angular rate as the Earth's rotation. A satellite in this orbit has a period of 24 hours and would hover over the same point on the equator. Equation 5 provides the relations for orbital period:

$$\text{Period} = 2\pi\,\text{Sqrt}(a^3/\mu) \tag{5}$$

From this, we can compute a geosynchronous orbit's radius to be 42,241 km. μ is the Earth's gravitational parameter, 398,601.2 km^3/sec^2.

Elliptical orbit. Elliptical orbits are much more difficult to analyze because the orbital velocity is not uniform. The satellite moves the fastest near perigee and the slowest near apogee. This is easily observed from Equation 6 (an orbit's energy and the fact that energy is conserved):

$$\text{Energy} = v^2/2 - \mu/r, \tag{6}$$

where r and v are the orbit's instantaneous radius and velocity. When the orbital velocity's nonuniformity is combined with nodal and perigee precessions and the Earth's rotation, the coverage analysis becomes extremely difficult. However, the motion's nonuniformity implies that the satellite will linger over some regions and rush through others. If the J_2 effects can be controlled, the coverage's nonuniformity can be exploited to great advantage. The next class of orbits provides the control.

What makes the elliptical orbit particularly difficult to handle under J_2 perturbation is the perigee's rotation. From Equation 4, we note that when the inclination is ±63.4 deg, the perigee is fixed. This special inclination is called critical inclination. Critically inclined elliptical orbits are very useful. The most famous of these is the Molniya orbit, which is a highly elliptical 12-hour-period orbit the Soviets originally designed to observe the northern hemisphere. A typical design sets the perigee at 200-km alti-
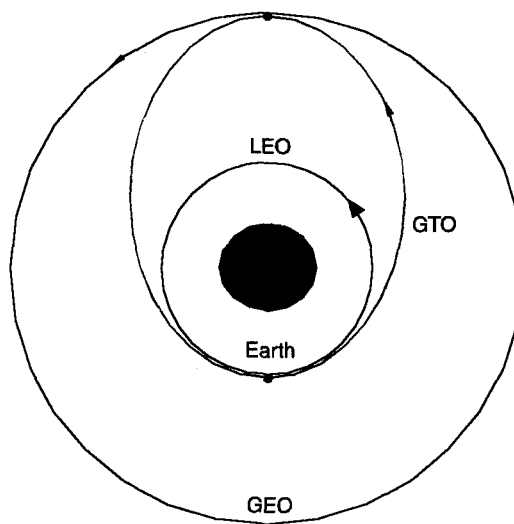

Figure 4. The GTO-Hohmann transfer of a satellite from a circular LEO to the high-altitude GEO.

tude, with an argument of perigee at −90 deg (over the south pole). Critical inclination is selected to fix the perigee over the south pole and maintain the observation geometry over the northern hemisphere, where this orbit spends most of its time. Figure 5 plots the Molniya orbit's groundtrack. Note how the apogee is over both the continental US and Russia. This permits the satellite to gather data over the US and transmit it while over Russia.

Periodic groundtrack orbit. The geosynchronous orbit and the Molniya orbit both have periodic groundtracks. For the geosynchronous orbit, the groundtrack is just a point. The Molniya orbit has a U-shaped one. For planetary observations, scientists often must take data over the same ground locations repetitively under varying conditions. It is possible to balance nodal regression with the Earth's rotation rate to ensure the groundtrack repeats after several orbits.

Sun-synchronous orbit. To maintain the same lighting conditions for observing a planet, it might be necessary to maintain a constant geometry between the sun and the satellite. We can achieve this by adjusting the nodal regression rate that precesses the orbital plane to match the sun's one deg/day motion. Such an orbit will always visit the same latitude at roughly the same time of day, thereby providing a near-constant lighting condition to observe the same location. Figure 6a shows this geometry. In Figure 6b, we see the orbit from the sun's point of view, which shows that the Earth never occults this orbit.

This might be very useful if, for example, the

Figure 5. Molniya orbit's groundtrack.

mission requires a lot of power and the solar panels must always point at the sun. This geometry is maintained throughout the year. Perturbations due to the Earth's orbit ellipticity will cause the geometry to shift slightly if it is not compensated by maneuvers. This is generally acceptable for most missions. Besides providing constant power, the solar geometry's constancy also implies a more stable thermal environment that helps with hardware design and mission operations.

Sun-synchronous orbits tend to be near-polar orbits with inclination greater than 90 deg. From Equation 3, we see that for the nodal drift to be positive, the inclination must be greater than 90 deg. For typical Earth satellites using this orbit, the altitude is generally around 1,000 km, which yields a near-polar orbit with inclinations around 100 deg.

There are sun-synchronous orbits whose perigees are fixed due to judicious selections of the orbital eccentricity and balancing the effects of the J_2 and J_3 harmonics. These are known as frozen orbits. Frozen orbits for Earth further require $\omega = 90$ deg and for Mars, $\omega = -90$ deg.

Ellipso-type orbit. Until recently, constellation design has been restricted mostly to military or scientific applications. Civilian telecommunications have very different requirements and design drivers. In particular, for business applications, peak usage tends to occur between 9 am and 5 pm (the day side). Designers decided that an ideal situation would have the communications satellites bunch up over the day side and move quickly through the night side. This is achieved with a critically inclined, highly elliptical, sun-synchronous orbit with apogee placed near the desired location on Earth (such as New York City). The sun synchronicity ensures the satellite will always visit New York at about the same time of day. The high eccentricity enables

the satellite to linger over New York on the day side as long as possible and spend minimum time on Earth's night side. The critical inclination guarantees that the perigee and the apogee's location are both fixed.[1]

Basic constellation types

Now that I have gone through the orbital zoo, let's put these cells together and design a multicellular organism: the constellation. Recall that the chief reason why constellations are needed is that one satellite can cover only a limited portion of the Earth at any particular instant. For global coverage, whether for observation or communications, a constellation is the only solution. Fortunately, this solution fits in well with the current technological trend to decentralize control and distribute the process across a network. The key differences are that the constellation is a wireless network whose nodes are not fixed and that operates in a less accessible and more challenging environment.

One of the key drivers for current constellation design is the need for global wireless telephony services. Geosynchronous networks have been in service for some time, but these behemoth satellites are extremely costly to build, launch, and maintain. The current drive for smaller, cheaper satellites and lower cost has driven the network from a geosynchronous altitude of over 35,000 km down to around 1,000 km for mobile satellite constellations. This drop in altitude represents many cost differences: satellites are cheaper to launch into orbit, the telecom and power requirements might be reduced, and the overall satellite hardware is cheaper to build. But the downside is that now instead of three to 10 geosynchronous satellites, close to 1,000 satellites might be required to provide global coverage.

In principle, constellation zoology is much more complex because arbitrary combinations of orbits might be considered in building a constellation. But in practice, due to the enormous difficulty in analyzing constellation performance, a few have evolved and emerged as useful design concepts. I reiterate here the enormous difference and technical challenge a network presents to the analyst as opposed to the task of analyzing a single network node. The combinatorial complexity exhibited in the factorial growth of the number of branches in the analysis tree is a major factor for the increase in difficulty.

Another high-level driver in constellation design is the system's communications design. Although this is an obvious fact, its impact on the
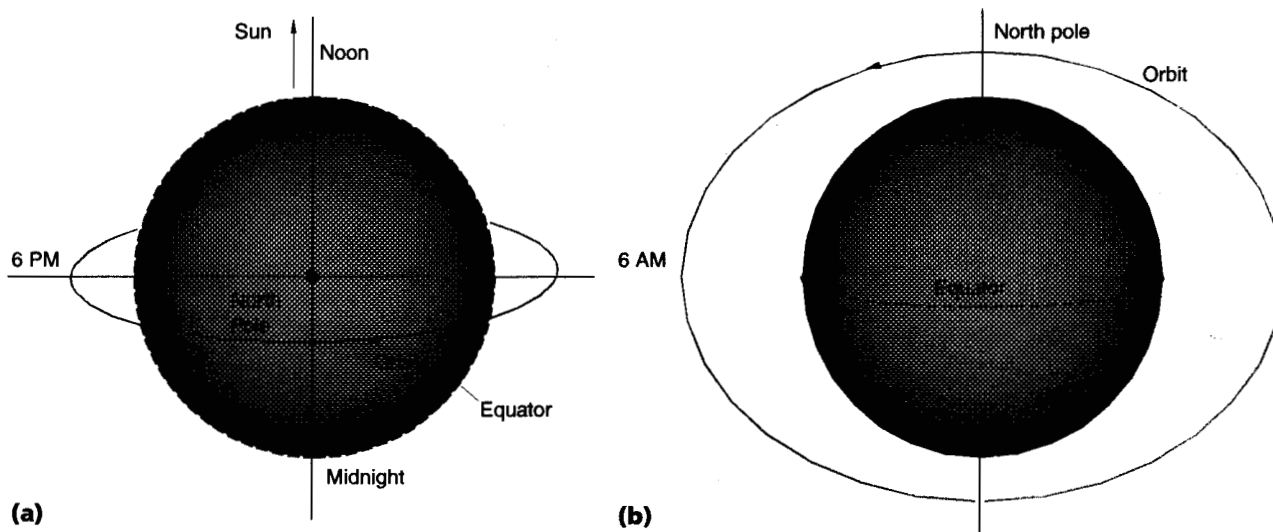
Figure 6. (a) 6 am view of sun-synchronous orbit from above the North Pole, and (b) 6 am view of sun-synchronous orbit from the sun's view.

constellation design is crucial and cannot be overemphasized. For example, whether the constellation's satellites have intersatellite communications links can make a big difference in the number of satellites the constellation requires. This, in turn, has a large impact on total system cost. Unfortunately, system trades of this type cannot be fully addressed in this limited article.

Constellations with circular orbits

Circular orbits are well-understood and provide the most uniform and stable design component for a constellation. In the case where 24-hour uniform global coverage is desired, a constellation with circular orbits provides an optimal solution.

Geosynchronous constellations. From Equation 1, we can estimate a geosynchronous satellite's coverage circle to have an 81-deg half-cone angle. This implies that with three geosynchronous satellites, we can cover between 81- and –81-deg latitude of the entire world geometrically. The NASA TDRSS system used this design to provide a global communications network for NASA missions. From the spacecraft and ground communications systems design, more spacecraft might be required to provide service with higher quality or greater throughput capacity. In general, such a system would require at most tens of satellites as opposed to the thousand-satellite system using LEOs or MEOs.

Walker constellation. Suppose you wished to design a communications network on the ground. A first attempt might be to uniformly divide the region with a Cartesian grid and place a communications node at each of the grid points. On the sphere, aside from the regular solid polygons well-known to the Greeks, there is no simple way to divide the sphere into regular Cartesian grids. Even if this were possible, there is no means at present to maintain a network of satellites at fixed nodes. Mathematically, this is the challenging geometric combinatorial problem of tiling on the sphere.

The next best thing is to preserve as much symmetry as possible by using the same template circular orbit repeatedly in a regular and sensible fashion. Three parameters describe the Walker constellation:[2] *T/P/F. T* is the total number of satellites, *P* is the number of orbit planes, and *F* is the phasing parameter. The *T* satellites are equally divided among *P* planes with the same inclination. The planes are evenly spaced by 360/*P* deg. The phase-angle offset is given by 360 *F/T* deg to ensure a more optimal packing of Earth's coverage circles. For example, Motorola's Iridium constellation design is based on the Walker constellation.

Streets-of-coverage constellation. Streets of coverage approaches the problem slightly differently from the Walker algorithm.[3,4] The basic idea is to design the coverage for a single plane of orbits first, then extend it to the entire globe by duplicating the satellite plane repeatedly. Start with an orbit plane and populate it with *K* satellites evenly spaced with overlapping coverage circles as shown in Fig-
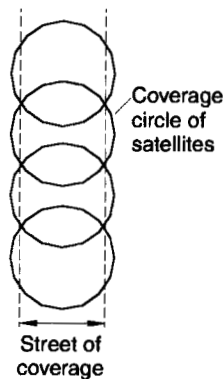
ure 7. This ensures the K satellites around the orbital plane cover an entire band. This band is called a street of coverage. The K-satellite plane is duplicated multiple times until various streets of coverage cover the equatorial region. This guarantees an annular region between two latitudes (determined by the inclination and altitude) is completely covered.

Figure 7. Street of coverage.

Constellations with elliptical orbits

When nonuniform coverage is required, elliptical orbits are much more efficient. The Molniya orbit is a good example because it is designed to spend most of its 12-hour period over the northern hemisphere. In fact, it was designed to gather information over the US for 12 hours and then relay the data over the Soviet Union during the next 12-hour orbit. Thus if one requires 24-hour data gathering, at least two satellites are needed. For redundancy and smoother transitions between satellites taking data, a three-or-more satellite constellation in Molniya orbit is desirable. They can be designed to have the same groundtrack but observe the same locations at different times of day.

With a colleague, I designed the Nuonce constellation (Nonuniform Optimal Network Communications Engine, using elliptical LEOs) specifically to target the western hemisphere during business hours.[1] The constellation used critically inclined, sun-synchronous, elliptical orbits with the apogee over the northern hemisphere. Recall that sun-synchronous orbits always cross the same latitudes during the same time of day. The Nuonce constellation concentrated the orbit planes that cross New York City between 9 am and 11 am, and 1 pm to 5 pm. This concentrates the coverage during the busiest times over business hours where peak support is needed. This independently developed constellation concept resembles Draim's copyrighted Ellipso orbit.

Elliptical orbits are generally more difficult to handle. In addition to the problem with the argument of perigee precession, the perigee passage itself is problematical due to drag effects that cause the orbit to decay or lose its desirable characteristics. Of course, these can be corrected with maneuvers. But when you have a 1,000-satellite constellation, operational simplicity becomes a critical cost issue—any means to avoid maneu-

vers should be carefully considered. Raising the perigee while keeping the semimajor axis (hence period) constant solves the decay problem but greatly lessens the elliptical orbit coverage strategy's effectiveness. Consequently, most designers tend to shy away from elliptical orbits, especially for LEO constellations where the drag problem can be very serious indeed.

Constellation performance metric

The two most critical satellite functions are observation and communication. In both instances, the service-frequency availability is the crucial metric for performance. Of course, the word *available* is loaded in that it depends highly on the nature of the service required and the hardware involved. An infrared detector is available to observe a patch of dark sky only when it is pointed away from the sun, Earth, or moon, but a camera photographing the Earth must view it in sunlight. The most primitive metric is geometric coverage: how often is the region in the spacecraft's line of sight and available for observation by some instrument? Even this simplified metric is not trivial for a single satellite.

An application of ergodic theory

Let us consider a classical problem: how often does a satellite see a ground station? A view period is the amount of time a satellite is visible to a ground station during a flyover. Of course, integrating the orbit and accumulating the times when the satellite sees the ground station easily simulates this. When there are multiple satellites and ground stations, once again the combinations overwhelm the situation even with simple propagators.

Instead of attaching the view circle to a satellite groundtrack's nadir point, one can attach the view circle to the ground station. Visibility between satellite and ground station is achieved only if the satellite nadir crosses into the station-view circle as in Figure 7. This picture brings to mind the Poincare Recurrent Theorem, which states that given a point x wandering ergodically in a box of volume 1, the probability of finding x in region A in the box is equal to A's volume. Put another way, the percent of time x spends in A equals Volume(A) × 100. If this holds for the satellite view-period problem, then the station-view-circle's area divided by the area the satellite groundtrack band (Figure 8) covers should provide coverage probability, hence the long-term average of the coverage view period. Un-

fortunately, the satellite groundtrack's motion on Earth is not ergodic—the Recurrence Theorem does not apply.

However, the reason it fails is due to the fact that satellite groundtracks do not move uniformly across Earth. They tend to bunch up near the poles. What if you could spread them apart as you go up in latitude closer to the pole? Indeed, using a weight function to compute the area can do this. Mathematically, this is finding a measure (weight function) that is invariant under the groundtrack's flow. In other words, an area weighted with the measure remains constant under the flow of the groundtrack. I found such a measure for circular orbits and derived an integral that provides the probability for a satellite to be in a given ground station's view.[5] This integral replaces the differential-equation dynamics required to generate the station-view periods needed to compute conventional algorithm probability.

When using this approach, it becomes practical to ask global questions such as, given all satellites in a circular orbit with a 1,000-km altitude, what are the best locations to place a ground station for maximum contact time between the satellite and the station? From this theory, we see that stations at latitudes near the poles tend to provide the most contact time (because the satellite groundtracks tend to bunch up near the poles). Figure 8 shows the view circles of two stations, one near the equator and one near the satellite footprint band's upper edge. The surprising fact is that both stations have the same probability of seeing the satellite even though the view-circle intersections with the band appear different under the standard sphere measure.

Visual calculus

The ergodic-theory application suggests that a geometric approach to coverage computation might be useful. Typically, coverage analysis is produced by first discretizing the sphere into a set of polygons. The propagated groundtrack is checked against the polygons to determine whether it is inside or outside, and statistics are accumulated to generate coverage history. Sphere discretization is not a trivial matter—a uniform finite partition is hard to come by because there are only five known classical solids. Aside from the problem of how to uniformly partition the sphere, the task of keeping track of the statistics can quickly become a computational nightmare.

An answer to this problem is visual calculus (based on the fact that an infinitesimal polygon provides a uniform discretization of the sphere).
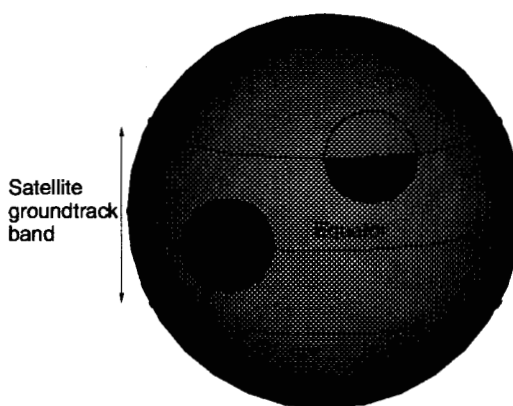

Figure 8. Two stations with the same amount of view periods.

Suppose we use an equal-area project (such as the Peter's projection of the sphere) onto a rectangular grid. Now every point in the rectangle has the same infinitesimal area. On the computer, we might consider a pixel as an infinitesimal. Thus, if we generate a plot showing where the satellite groundtrack passed through the station-view circle, just by computing the number of pixels in the view circle under an equal area projection allows us to compute the coverage quickly. Furthermore, the map projections handle the data-organization problem. Once we compute maps with coverage information, they can be operated upon algebraically to produce further statistical products.

One limitation to this approach is that an estimate's accuracy depends on the map's resolution. To overcome this drawback, a multiresolution map function replaces the pixel-based map. This also separates the computation completely from the visualization. Nevertheless, due to its conception, we still call this method *multiresolution visual calculus*, which is currently under patent review at JPL.

Global coverage analysis

The reason for ergodic method development and visual calculus is to facilitate global-coverage-information computation. In the case of ergodic theory application, it enables a quick, global calculation of the station-view-period performance so that a planner can easily decide the best location to place a ground station based on knowledge of the satellites it is supposed to service. Prior to the ergodic method, such a computation would have been nearly prohibitive due to the large amounts of computation required.
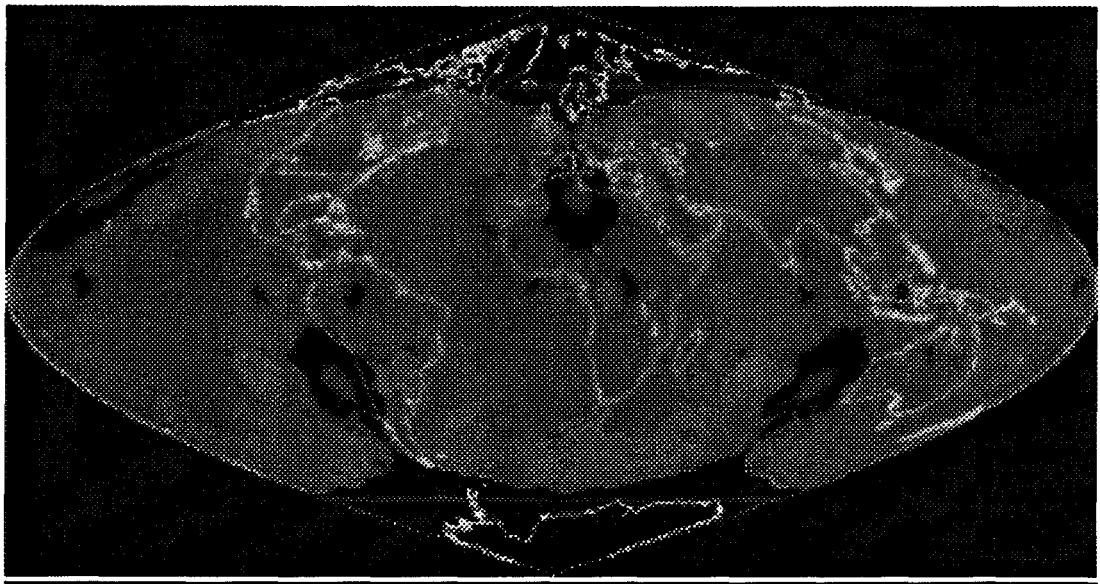
Figure 9. Revisit-time map of a double constellation with radar instruments.

Similarly, the visual calculus is designed to provide global statistics of visibility, revisit, and other coverage data. In Figure 9, I provide a revisit-time map computed using visual calculus for a complex mission. In this instance, two constellations each carry different instruments. The two constellations are at different altitudes to optimize their respective instrument performance. The coverage analysis requires modeling instrument performance, which is highly sensitive to geometry. Visual calculus provides a language and algorithm to compute coverage statistics to answer the revisit question: how long do I have to wait until the next satellite can see me? In this instance, the green area provides a revisit time of less than 30 minutes, the yellow area provides a revisit time of less than one hour, and the orange area provides a revisit time greater than one hour. The instruments do not view the black region at all. Note the resulting map's complex geometry.

Exotic constellations

Two exotic constellation concepts are under consideration for various missions. The first concept is formation flying. The idea is to put a group of satellites in orbit around Earth and force them to fly in a geometric pattern such as a triangle. In the case of loose formations, this is reasonable. The polygon formation is not required to be rigid and might flex and change shape as the orbits evolve. Recall that on a sphere, two great circles must always intersect at the antipodes. This fact forces two circular orbits of the same radius to always intersect twice per orbit. For elliptical orbits, the intersection patterns might be more complex, but it is virtually impossible to enforce a rigid polygon formation without propulsion and control. The cost of maintaining a rigid formation for long periods is prohibitive using the brute-force approach. The only viable ones are the string-of-pearls formation, where a series of satellites follow one another in the same orbit. Another formation concept is to use heliocentric orbits (such as the Earth's orbit), where gravitation is much more uniform. In this case, a rigid formation might be held together for some time because the orbit changes so slowly.

The second exotic concept is a quasihalo constellation (Figure 10) around the L_1 Lagrange point. L_1 is a point between the sun and the Earth, roughly 1.5 million km away from Earth along the sun-Earth line (where Earth/sun gravitation balances with the rotational force to produce a relative equilibrium). A particle placed at L_1 will remain there forever, if no perturbations are introduced. But L_1 is unstable, and the region around it in the phase space is chaotic. There are 3D quasihalo orbits around L_1 that wind around a torus.[6,7] These orbits provide the dynamics to control a constellation near L_1 flying in formation.

The design of such constellations requires an understanding of dynamical systems theory, particularly of invariant manifolds. A great deal of work is required to understand the dynamics of these orbits and how to control them. In partic-

ular, continuous thrusting is required for their control, perhaps with ion engines. This requires an understanding of optimal control theory in the three-body problem—currently an active area of research and development.

Satellite constellations are a part of our daily lives, from DirectTV to our cell phones and car navigation—it's a modern fact of life. Currently a network of geosynchronous satellites provides our global communications infrastructure. Preliminary results for elliptical orbits have been established but are incomplete. The problem of multiple satellites has not been investigated.

Although satellite constellations have been around for some time, the recent interest in the satcom industry has given the subject new life because so much is at stake. But despite the fact that this is a multibillion-dollar industry just waiting to take off, the efforts expended in the area of mission analysis and design are disproportionately small. Considering how much cheaper simulation and analysis are compared with attempts to fix complex system problems after launch, this is incomprehensible.

This problem is a result of the past success of the satellite industry, where so much has been done with seemingly so little investment into mission design and trajectory analysis. The growth of technology has created much more complex instruments and even more complex mission requirements unimagined in the 1950s. Despite the incredible advances in computation hardware and software, at the end of the 20th century, we have reached a plateau where the old trajectory technology no longer adequately serves modern mission requirements. New methods, new paradigms such as dynamical systems theory, chaotic orbits, ergodic theory, and so forth, will provide new capabilities to respond to new challenges. 🔊
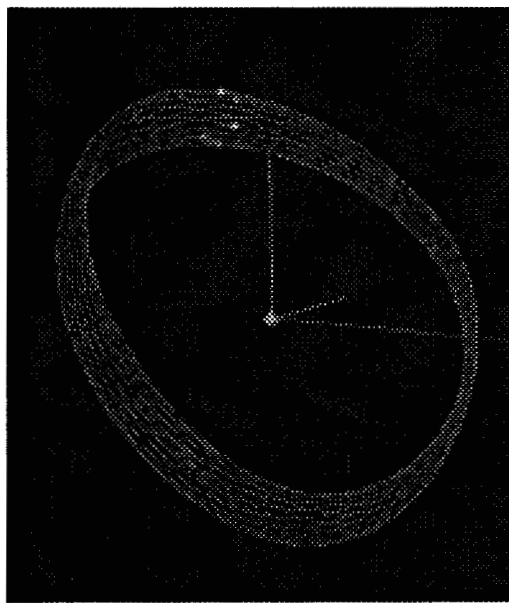
Figure 10. A quasihalo constellation near L1.

References

1. M. Lo and P. Estabrook, "The NUONCE Engine for LEO Networks," *Proc. Fourth Int'l Mobile Satellite Conf.*, 1995, pp. 193–197.

2. J.G. Walker, *Circular Orbit Patterns Providing Whole Earth Coverage*, Tech. Report 70211, Royal Aircraft Establishment, 1970.

3. R. Luders, "Satellite Networks for Continuous Zonal Coverage," *American Rocket Soc. J.*, Vol. 31, Feb. 1961, pp. 179–184.

4. W. Adams and L. Rider, "Circular Polar Constellations Providing Continuous Single or Multiple Coverage above a Specific Latitude," *J. Astronautical Science*, Vol. 35, No. 2, Apr.–June 1987, pp. 155–192.

5. M. Lo, "The Invariant Measure for the Satellite Ground Station View Period Problem," *Hamiltonian Systems with Three or More Degrees of Freedom*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1998.

6. G. Gomez et al., "Lissajous Trajectories Around Halo Orbits," Tech. Report AAS 97-106, AAS/AIAA Space Flight Mechanics Meeting, Huntsville, Ala., Feb. 1997.

7. K. Howell and B. Barden, "Formation Flying in the Vicinity of Libration Point Orbits," Tech. Report AAS 98-169, AAS/AIAA Space Flight Mechanics Meeting, Monterey, Calif., 1998.

Martin W. Lo is a member of the Navigation and Flight Mechanics Section at the Jet Propulsion Laboratory. Lo received his PhD from Cornell University and BS from Caltech. His technical interests include trajectory design and dynamical systems theory. Contact him at Jet Propulsion Laboratory, 4800 Oak Grove Dr., Pasadena, Calif. 91109; mwl@trantor.jpl.nasa.gov.